

I. Dérivées

1. Ligne de niveau

$$N_c = \{x \in \mathbb{R}^n | J(x) = c\}$$

2. Dérivée première

a. Gradient

$$\nabla J(x_0) = \left[\frac{\partial J}{\partial x_1}, \dots, \frac{\partial J}{\partial x_n} \right]^T$$

Propriété : Au point x_0 , $\nabla J(x_0)$ est $\perp$ à la ligne de niveau, son sens va dans le sens de J croissant.

b. Dérivée directionnelle

$$D_x J(x, d) = \lim_{\varepsilon \rightarrow 0} \frac{J(x + \varepsilon d) - J(x)}{\varepsilon} = \left. \frac{d \overbrace{J(x + \varepsilon d)}^{\varphi(\varepsilon)}}{d\varepsilon} \right|_{\varepsilon=0} = \varphi'(0) = \nabla J(x)^T d$$

c. Plan tangent

Passe par $(x_0, J(x_0))$ et vecteur directeur $\nabla J(x_0)$

Approximation locale de J

$$P = \{x \in \mathbb{R}^n | \nabla J(x_0)^T (x - x_0) = 0\}$$

$$J(x) \approx P(x) = J(x_0) + \nabla J(x_0)^T x$$

d. Développement limité au premier ordre

$$J(x) \approx J(x + \varepsilon d) = J(x) + \varepsilon \underbrace{\nabla J(x)^T d}_{DJ_x(x, d)} + o(\varepsilon)$$

e. Règles de calcul pour la dérivée

$$D(J_1 + \alpha J_2) = DJ_1 + \alpha DJ_2 \quad | \quad D(J_1(J_2)) = D_z J_1(z = J_2) D_x J_2(x) \quad | \quad \nabla a^T x = a \quad | \quad \nabla x^T A x = 2Ax$$

3. Dérivées secondes

a. Dérivée directionnelle au sens de Gâteaux

$$D^2 J(x, d) = \lim_{\varepsilon \rightarrow 0} \frac{DJ_x(x + \varepsilon d) - DJ_x(x)}{\varepsilon}$$

b. Matrice Hessienne

$$H_J(x) = \begin{pmatrix} \frac{\partial^2 J(x)}{\partial x_1 \partial x_1} & \dots & \frac{\partial^2 J(x)}{\partial x_1 \partial x_j} & \dots & \frac{\partial^2 J(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial^2 J(x)}{\partial x_i \partial x_1} & \dots & \frac{\partial^2 J(x)}{\partial x_i \partial x_j} & \dots & \frac{\partial^2 J(x)}{\partial x_i \partial x_n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial^2 J(x)}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 J(x)}{\partial x_n \partial x_j} & \dots & \frac{\partial^2 J(x)}{\partial x_n \partial x_n} \end{pmatrix}$$

Calcul pratique :

A partir de la dérivée de

$$\varphi(\varepsilon) = D_x J(x + \varepsilon d)$$

Identifier H dans $\varphi'(0)$:

$$\varphi'(0) = d^T H_X(x)^T d$$

c. Développement limité au second ordre

$$J(x + d) = J(x) + D_x J(x, d) + \frac{1}{2} D_x^2 J(x, d) + o(\|d\|^2)$$

$$J(x + d) = J(x) + \nabla_x J(x)^T d + \frac{1}{2} d^T H d + o(\|d\|^2)$$

II. Généralités

Variables V	$V \in \mathbb{R}^p$
Objectifs O	$V \mapsto \mathbb{R}^q \quad \min_{x \in \Omega} J(x)$
Contraintes C	$V \mapsto \mathbb{R}^{m+n} \quad \begin{cases} h_i(x) = 0 \\ g_j(x) \leq 0 \end{cases} \Leftrightarrow \begin{cases} H(x) = (h_1, \dots, h_n)^\top = 0 \\ G(x) = (g_1, \dots, g_m)^\top \leq 0 \end{cases}$
Domaine de faisabilité Ω	$\Omega = \{x \in \mathbb{R}^d ; H(x) = 0 \text{ et } G(x) \leq 0\}$
Fonction coût	$J: \Omega \mapsto \mathbb{R}$
Domaine de la fonction coût	$\text{dom } J = \{x \in \Omega ; -\infty < J(x) < +\infty\}$ Si $\emptyset$, fonction coût impropre : pas de solution
Minimum global θ^*	$J(\theta^*) \leq J(\theta) \quad \forall \theta$
Minimum local $\hat{\theta}$	$J(\hat{\theta}) \leq J(\theta) \quad \forall \theta \mid \ \hat{\theta} - \theta\ \leq \varepsilon$

III. Optimisation convexe sous contraintes

1. Avec contraintes d'égalités

Problème	Lagrangien	Conditions d'optimalité
$\begin{cases} \min_{x \in \mathbb{R}^n} J(x) \\ \text{s. c. } H(x) = 0 \end{cases}$	$L(x, \lambda) = J(x) + \sum_{j=1}^p \lambda_j H_j(x)$ λ_j multiplicateurs de Lagrange	$\nabla_x L = 0 \Rightarrow \nabla_x J(x) + \sum_{j=1}^p \lambda_j \nabla_x H_j(x) = 0$ $\nabla_{\lambda_j} L = 0 \Rightarrow H_j(x) = 0$

2. Avec contraintes d'inégalités

a. Lagrangien et conditions d'optimalité

Problème	Lagrangien	Conditions d'optimalité KKT
$\begin{cases} \min_{x \in \mathbb{R}^n} J(x) \\ \text{s.c. } H(x) = 0 \\ \text{et } G(x) \leq 0 \end{cases}$	$L(x, \lambda, \mu) = J(x) + \sum_{j=1}^p \lambda_j H_j(x) + \sum_{i=1}^q \mu_i G_i(x)$ λ_j multiplicateurs de Lagrange	Stationarité : $\nabla L(x, \lambda) = 0$ Adm. primale : $G(x) \leq 0$ Adm. duale : $\mu_i \geq 0$ Complémentarité : $\mu_i G_i(x) = 0$

b. Problème dual

Le problème dual est $\mathcal{L}(\mu, \lambda) = \min_x L(x, \lambda, \mu)$

IV. Optimisation convexe sans contraintes

1. Matrice définie positive

A définie positive	A définie négative	A quadratique
$\lambda_i \leq 0$ J convexe	$\lambda_i \geq 0$	$\begin{cases} \lambda_i \leq 0 \\ \lambda_j \geq 0 \end{cases}$

2. Condition d'optimalité

a. Existence de solution

J est coercivité	$\ x\ \rightarrow \infty \Rightarrow J(x) \rightarrow \infty$ (fonction infinie à l'infini)
J est propre	$\exists x \mid J(x) \in \mathbb{R}$
$\exists$ solution globale	Si J continue, propre, coercive, $\min_x J(x)$ admet une solution globale

b. Conditions d'optimalité

1^{er} ordre	x_0 solution $\Rightarrow \nabla J(x_0) = 0$
2^{ème} ordre	x_0 solution $\Rightarrow \nabla J(x_0) = 0$ et $H(x_0)$ définie positive $\nabla J(x_0) = 0$ et $H(x_0)$ définie positive $\Rightarrow x_0$ solution locale
Convexité	J est convexe et x_0 respecte condition d'optimalité $\Rightarrow x_0 = x^*$ solution globale

3. Optimisation itérative

a. Recherche linéaire (line search)

$$x_{k+1} = x_k + \rho_k d_k \quad \lim_{k \rightarrow \infty} x_k = x^* = \operatorname{argmin}_{x \in \mathbb{R}^n} J(x)$$

i. Direction de descente d ($\nabla J^T d < 0$)

- **Gradient** : $d_k = -\nabla J$ $\mathcal{O}(n)$
- **Gradient conjugué** : $d_k = -\nabla J + \beta_k d_{k-1}$ $\mathcal{O}(n^2)$
- **Quasi-Newton** : $d_k = -B \nabla J$ (Ex : $B = \operatorname{diag}(H)^{-1}$) $\mathcal{O}(n^2)$
- **Newton** : $d_k = -H^{-1} \nabla J$ $\mathcal{O}(n^3)$

ii. Choix du pas p

- **Pas fixe** : $\rho_k = c$ $\mathcal{O}(1)$
- **Pas variable** : $\rho_k = \begin{cases} \alpha \rho_{k-1} & \text{si } J \text{ diminue} \\ \beta \rho_{k-1} & \text{si } J \text{ augmente} \end{cases}$ $\alpha = 1,15$ $\beta = 0,5$ $\mathcal{O}(1)$
- **Pas optimal** : $\rho_k = \begin{cases} \operatorname{argmin}_{\rho \in \mathbb{R}^+} J(x + \rho d) \\ \frac{d^T d}{d^T H d} \end{cases}$ $< \mathcal{O}(n^2)$ $\mathcal{O}(n^2)$

b. Région de confiance (trust region)

Comme recherche linéaire avec saut maximum limité par une région de confiance

$$x_{k+1} = x_k + d_k \quad \text{avec} \quad d_k = \operatorname{argmin}_{d \in R(x_k)} M(x_k + d)$$

$R(x)$: région de confiance autour de x $M(x + d)$: modèle dans la région de confiance (DL)

V. Programmation linéaire

1. Forme standard

$$\begin{cases} \min_{z \in \mathbb{R}^m} c^T z \\ Az = b \\ z \geq 0 \end{cases} \quad \begin{array}{l} z : m \text{ inconnues} \\ c : m \text{ couts} \\ A : p \text{ contraintes } (p < m) \\ b : p \text{ seconds membres des contraintes} \end{array}$$

2. Méthodes de réduction à la forme standard

a. Inégalités $\Rightarrow$ variables d'écart (positives)

$$z = \begin{bmatrix} x \\ e \end{bmatrix} \quad \begin{array}{l} f_i(x) \geq b_i \Leftrightarrow f_i(x) - e_i = b_i \\ f_i(x) \leq b_i \Leftrightarrow f_i(x) + e_i = b_i \end{array}$$

b. Variables non contraintes $\Rightarrow$ décomposition

$$x_i \text{ sans contraintes remplacé par } \begin{cases} x_i = x_i^+ - x_i^- \\ x_i^+, x_i^- \geq 0 \end{cases}$$

3. Forme duale standard

$$\begin{cases} \max_{y \in \mathbb{R}^p} b^T y \\ A^T y \leq c \end{cases} \quad c^T x^* = b^T y^* \quad y = \lambda$$

4. Méthode du point intérieur

a. Principe

Résolution d'un problème à la forme standard en résolvant simultanément le primal et le dual à travers les KKT, en les écrivant sous la forme $F(z, \lambda, \mu) = 0$ avec $z, \mu \geq 0$.

C'est une méthode itérative de Newton projeté appliquée à $F \begin{pmatrix} \Delta z \\ \Delta \lambda \\ \Delta \mu \end{pmatrix}$.

b. Maths et algorithme

$$\begin{cases} \min_{z \in \mathbb{R}^m} c^T z \\ Az = b \\ z \geq 0 \end{cases} \quad \begin{cases} A^T \lambda + \mu - c = 0 \\ Az = b \\ \text{diag}(\mu)z = 0 \\ z \geq 0 \\ \mu \geq 0 \end{cases} \Leftrightarrow F(z, \lambda, \mu) = 0$$

$$\begin{pmatrix} z \\ \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} z \\ \lambda \\ \mu \end{pmatrix} + \underbrace{\text{pas}}_d \times \begin{pmatrix} \Delta z \\ \Delta \lambda \\ \Delta \mu \end{pmatrix}$$

Avec un DL au 1^{er} ordre :

$$\underbrace{F(x+d)}_{=0} = F(x) + d^T \nabla_x F(x) + \underbrace{o(\|d\|)}_{=0} \Rightarrow d = \nabla_x F(x)^{-1} F(x)$$

$$\nabla_x F(x) = \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ \text{diag}(\mu) & 0 & \text{diag}(z) \end{pmatrix}$$

Forme matricielle de F :

$$F(z) = \begin{matrix} & m & p & m \end{matrix} A_F \times \begin{matrix} m \\ p \\ m \end{matrix} x_F + b_F$$

$$\begin{matrix} m \\ p \\ m \end{matrix} \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ \text{diag}(\mu) & 0 & 0 \end{pmatrix} \times \begin{pmatrix} z \\ \lambda \\ \mu \end{pmatrix} + \begin{pmatrix} -c \\ -b \\ \delta\sigma \end{pmatrix}$$

Pour que ça marche, on met $\delta\sigma$ dans b_F avec $\delta = \mu^T z$ et $\sigma \ll 1$

On calcule le pas tel que $z \geq 0, \mu \geq 0$.